Object Tracking in Video Sequence

Mr.Abdul Azeem[1], K.V.Sai Sireesha[2], A.Reshma Bhavani[3], K.Divya[4]
[1]*Assistant professor,* [1234]*Department of ECE*
[1234]*Andhra Loyola Institute of Engineering And Technology*
Vijayawada, Andhra Pradesh, India

Abstract- Today, there is a number of applications based on image processing and video processing. For instance, in surveillance systems, in security systems, etc., Object detection and tracking in real time is a complex area of research in computer vision. Object tracking definition is to track the required object over a sequence of images. There are mainly three steps in video analysis, detection of the required object, tracking of that object from frame to frame, and analysis of the tracked object to recognize their behavior. In this paper, we are trying to track the required object continuously in a video sequence. To track the object we selected TLD(Tracking-Learning-Detection)framework which is used for long term tracking. For feature extraction of the object, which we want to track, we apply SIFT(Shift Invariance Feature Transform). By using both TLD and SIFT algorithms we tracked the object. After feature extraction of the object the tracker follows the object from frame to frame. The detector will corrects the error if any error occurred in the tracker. The learning can identify the detector's errors.

Keywords- video processing, computer vision, TLD, SIFT.

I. INTRODUCTION

Object tracking in a video sequence, is one of the important research areas due to its commercial applications such as surveillance systems, security systems, mobile robots. The problem and difficulty to track the object depend on several factors such as the amount of important knowledge about the target and the number of type of parameters being tracked (scale, area, weight, rotation angle, shape).

Based on the classification of object tracking, video tracking can be divided into single target video tracking and multi-target video tracking. In this paper, we are going to track a single target. The single target tracking mainly consists of three steps, detection of required moving object, tracking of that object from frame to frame, and analysis of the tracked object to observe their behavior.

To get the long term tracking, there are a number of problems which need to be rectified. The main problem is the detection of the object when it reappears in the video, next a successful long-term tracker should handle the scale and illumination changes, background clutter, occlusions. The long-term tracking can be achieved either from tracking or from detection perspectives. Tracking algorithms identify object motion. Trackers require initialization, are fast and it produces smooth trajectories. On the other hand, trackers accumulate error during run-time and it fails if the object disappears from the video. Detection algorithms estimate the object location in every frame independently. Detectors do not mound and do not fail if the object disappears from the video.

The first contribution of this paper is the design of a TLD framework that divides the long-term tracking task into three sub-tasks: tracking, learning and detection. Each sub-task is addressed by a single component and these components operate simultaneously. In the TLD framework tracker follows the object from frame to frame. The detector detects the object and corrects the tracker if necessary. The learning expects the detector's error and updates it to avoid these errors in the future.

II. RELATED WORK

This section reviews the related approaches for each of the component of our system. Section 2.1 reviews object tracking. Section 2.2 reviews object detection.

A. OBJECT TRACKING

Object tracking is the task of estimation of the object motion. Trackers assume that object is visible throughout the sequence. Various representations of the object are used in practice, for example: points [1], [2], articulated models [3], contours [4], or optical flow [5]. In this paper we focus on the methods that represent the objects by geometric shapes and their motion is estimated between consecutive frames, i.e. so called frame-to-frame tracking. Template tracking is the most straightforward approach in frame-to-frame tracking. The object will be described by a target template.

B. OBJECT DETECTION

Object detection is the task of localization of objects in an input image. The object may be human, vehicle, missiles, etc. The definition of an object varies based on our requirement. It can be a single instance or a whole class of objects. The feature based approaches usually follow the pipeline of: (i) feature detection, (ii) feature recognition.

III. TLD FRAMEWORK

TLD is designed for long-term tracking of an unknown object in a video. TLD block diagram is shown in the below figure. Tracker expects objects motion between consecutive frames under the assumption that the object is visible. The trackers never recover if object moves out of video. Detector treats every frame as independent.

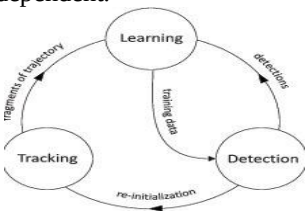

Fig.1: Block diagram of TLD framework

While a wide range of trackers and detectors exist but we are not aware of any learning method that would be suitable for the TLD framework.. The learning method[6] should: (i) deal with arbitrarily complex video streams where the tracking failures are frequent, (ii) never degrade the detector if the video does not contain relevant information and (iii) operate in real time.

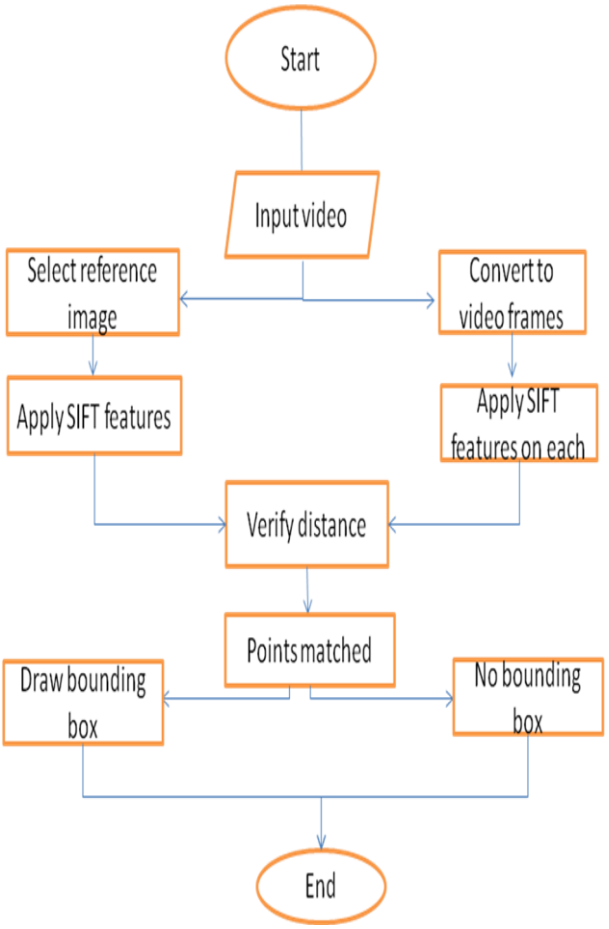

Fig.2: Flow chart of proposed algorithm

IV. SIFT ALGORITHM

SIFT stands for Scale Invariance Feature Transform. By using this algorithm we can extract the required features. Matching features across different images is a common problem in object tracking. When we have images of different scales and rotations, we need to use the Scale Invariant Feature Transform. SIFT is feature extraction algorithm in computer vision.

In this paper we selected scale, rotation angle and weights as the main features to detect the object.

The steps to extract the main features are as follows:

1. Let us consider

   L- Original image

   I- Gaussian filtered image

And take the difference between the L and I that should be considered as Difference of Gaussian.

DOG=L-I

2. Now apply the convolution between the original image and DOG

   K= DOG * I

   By using this we can extract some key points

3. After that do the localization of key points

   X= (DOG)□

   If I * X > 0.5 then that point should be considered as main key point

4. Now find out the magnitude and rotation angle

$$\sqrt{(L(x+1,y)-L(x-1,y))^2+(L(x,y+1)-L(x,y-1))^2}$$ this formula is used to calculate the magnitude and it is denoted as m(x,y).

$$\theta(x,y)=\tan^{-1}((L(x,y+1)-L(x,y-1))/(L(x+1,y)-L(x-1),y)))$$ this is formula is used to calculate the rotation angle.

5. If number matched features are greater than three then the bounding box formed.

6. Based on the matched features the object will be tracked and if more feature are matched then accuracy will be more.

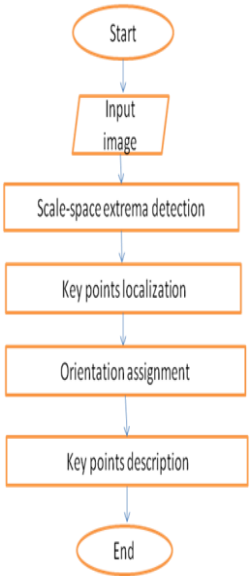

FIG.3: Flow chart for SIFT algorithm

V. RESULTS

The results for the object tracking in video sequence using TLD (Tracking-Learning-Detection) framework are as follows. Initially we have to select the interest object and by using SIFT algorithm features are calculated. After that the object is tracked.

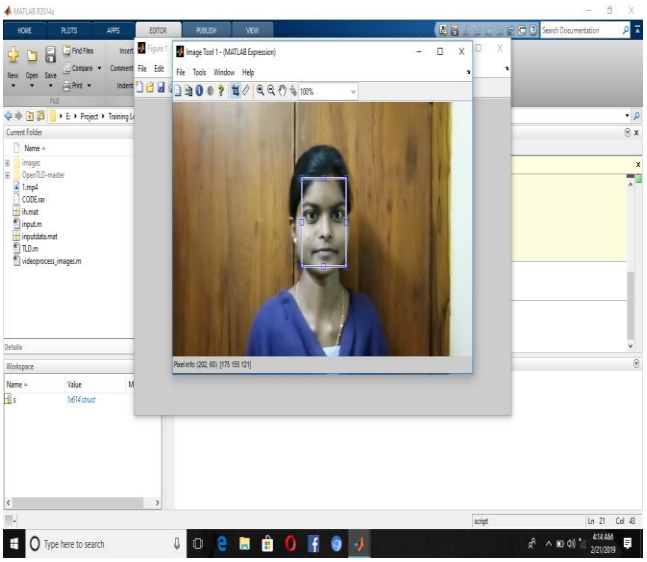

Fig.4: OBJECT SELECTION

Fig.5: TRACKING RESULTS

VI. CONCLUSION

In this paper, we studied the problem of tracking of an unknown object in a video stream. We designed a framework that divides the task into tracking, learning and detection.

First we have to select the required object from one frame. By using SIFT algorithm for feature extraction we extracted the required object features from reference image and by using that features of reference image we compare all frames with reference image. Based on the results of matched features the object will be tracked.

VII. REFERENCES

[1]. P. Sand and S. Teller, "Particle video: Long-range motion estimation using point trajectories," International Journal of Computer Vision, vol. 80, no. 1, pp. 72–91, 2008.

[2]. J. Shi and C. Tomasi, "Good features to track," Conference on Computer Vision and Pattern Recognition, 1994.

[3]. P. Buehler, M. Everingham, D. P. Huttenlocher, and A. Zisserman, "Long term arm and hand tracking for continuous sign language TV broadcasts," British Machine Vision Conference, 2008.

[4]. C. Bibby and I. Reid, "Robust real-time visual tracking using pixel-wise posteriors," European Conference on Computer Vision, 2008.

[5]. T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," European Conference on Computer Vision, pp. 25–36, 2004.

[6]. A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," Conference on Computational Learning Theory, p. 100, 1998.